Dependency2Vec

An alternate approach for word vector representations

Neural Embeddings

- NLP techniques for mapping words or phrases to dense vector representations
- Represent words in a continuous space, semantically similar words are mapped to nearby points
- All techniques depend on Distributional Hypothesis
- Enable efficient computation of semantic similarities of words and other linguistic relations

$$\boldsymbol{v}(\mathbf{king}) - \boldsymbol{v}(\mathbf{man}) + \boldsymbol{v}(\mathbf{woman}) = \boldsymbol{v}(\mathbf{queen})$$

- Techniques
  - word2vec (Mikolov et al., 2013 *Efficient Estimation of Word Representations in Vector Space*)
  - paragraph2vec (Le, Mikolov, 2014 *Distributed Representations of Sentences and Documents*)
  - GloVe (Pennington et al., 2014 *Global Vectors for Word Representation*)
  - **dependency2vec (Levy, Goldberg, 2014 *Dependency-Based Word Embeddings*)**
  - sense2vec (Trask et al., 2015 *A fast and accurate method for word sense disambiguation*)
  - FastText (Bojanowski et al., 2016 *Enriching Word Vectors with Subword Information*)

Context and Similarity

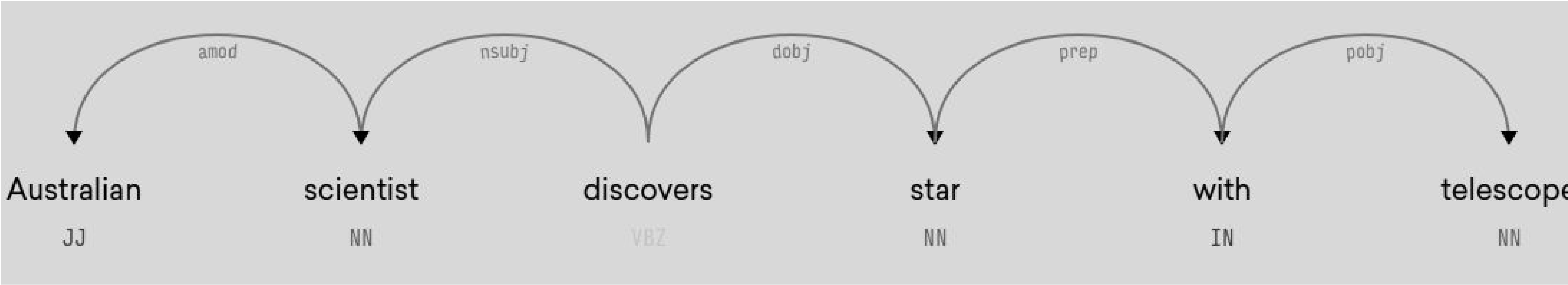
Topical Context (Relatedness)

- Embeds words by their domain similarity, words often seen together in sentences.
- Considers a context as a window of words surrounding a given target word.
- The context used by word embedding approaches like word2vec.
- Relies on a bag of words.
- Answers the question: what words are typically seen together in a sentence?

Australian scientist <u>discovers</u> star with telescope

Syntactic Dependencies

- Describes the process of analyzing the structure of a given sentence
- Tags each word with a part-of-speech (POS) tag that describes the word's syntactic function
- Determines the syntactic relationships between words (Dependency tags) in the sentence, represented with dependency parse tree.
- These syntactic relationships are directly related to the underlying meaning of the sentence in question.



Context and Similarity

Functional Context (Similarity)

- Embeds words their syntactic relationships in sentences.
- Considers a context as a target word, its dependency modifiers, and its root word.
- The context used by word embedding approaches like dependency2vec
- Answers the question: what words share the same functional meaning?

WORD	CONTEXT
australian	scientist/amod_INV
scientist	australian/amod, discovers/nsubj_INV
discovers	scientist/nsubj, star/dobj, telescope/prep_with
star	discovers/dobj_INV
telescope	discovers/prep_with_INV

dependency2vec

dependency2vec: SkipGram Model

- *word2vec* uses the *SkipGram* model to predict the surrounding window of context words from a target word
- Generates context:target pairs ([austrailian, discovers], scientist), ([scientist, star], discovers)....
- Inverts the pairs and attempts to predict each context from its target word.
- The task becomes to predict “austrailian” and “discovers” from “scientist”, “scientist”, “star” from “discovers”

dependency2vec: SkipGram Modification

- Where *dependency2vec* differs from *word2vec* is how it represents context.
- Is a modification of the *SkipGram* model to consider arbitrary contexts.
- No longer a linear window of a surrounding words, but the dependency context pairs we saw earlier.
- Both word embedding approaches operate in exactly the same way, the *SkipGram* context modification is the core difference.
- Advantages
  - Captures word relations that are outside of a linear window
  - Reduces the weighting of coincidental words, words in the same window that bare no useful relation.
- Intuition: words that appear in similar dependency contexts should have be functionally similar.

Comparing Model Ouput

word2vec vs dependency2vec

Target word: **Hogwarts**

word2vec	**dependency2vec**
harry	Castleobruxo
dubmbledore	Durmstrang Institute
magic	Harvard

word2vec vs dependency2vec

Target word: **Bitches**

word2vec	**dependency2vec**
Bitch - 0.868	Females - 0.866
Fuckers - 0.761	Hoes - 0.864
Fuck - 0.753	Mfs - 0.848
Asses - 0.734	Dudes - 0.812
Ass - 0.718	Niggaz - 0.806
Chicks - 0.711	Chicks - 0.783
Assholes - 0.710	Girls - 0.753
Fucking - 0.707	Boys - 0.750

Real numbers represent cosine similarity to the target word